USING HOME-MADE TESTS IN HIGH SCHOOLS

LEE BYRNE
University of Iowa
Formerly Supervisor of High-School Instruction, Dallas, Texas

This paper indicates briefly the ways in which some home-made tests have recently been employed in the high schools of Dallas, Texas. These are not published standardized tests but sets of questions prepared by the supervisor in the central office and based directly on the content of the local course of study.

The tests were not directed exclusively toward any one object. Rather it was the intention to secure some representative material from the different classes of the same grade in the various high schools and to examine this material with a view to ascertaining by interpretation whatever lessons it might prove to contain.

The different procedures employed and the kinds of inferences drawn are shown in the following pages. To the outside reader the most fruitful procedure will probably be the analyzing of the subject-matter into its subdivisions and the discrimination between the quality of results attained in the several subdivisions or phases of each subject. In every case it was revealed that notably higher results were attained in some aspects of the work than in others. In one instance it was discovered that an important phase of the subject was practically untaught (graphs in algebra). Such results pointed to the desirability of changes in school practice, some of which have already been carried out to good advantage.

THIRD-SEMESTER ALGEBRA

A general test was given to the twenty-three classes in third-semester algebra when they were about half-way through the term, and toward the end of the term they were given a second test on new topics.

Each paper received a simple mark, and the marks were tabulated. These marks, whose positions were rather accidental, de-

pending on the difficulty of the test and the method of marking, were then translated to a "standardized" or "Dallas scale," based on the probability curve and the assumption of 70 as the mark separating passing from failing, with a range of passing marks from minus $1\frac{1}{2}$ sigma to plus $2\frac{1}{2}$ sigma in the distribution. In other words, the upper four-fifths of the curve was assigned to passing marks and the lower fifth to failing marks. The relative position of each class, teacher, and school, as well as the city average, was indicated on each scale.

The averages in the two tests on a standardized scale were compared item by item with the teachers' marks for the first and second halves of the term and for the whole term. This was done by both tabulation and graphic drawing, leading to the inference that certain individual teachers were marking too high or too low as the case might be.

The Spearman correlation between the first and second tests was .48, between first half-term mark and second half-term mark .62, between term mark and the average of the two half-terms .81, between term mark and the average of the two tests .55, between average of the two half-terms and average of the two tests .51. A fair degree of reliability is thus indicated for all of the marks.

The distribution of correct answers was ascertained for each test, showing the number of pupils securing 8, 7, 6, 5, etc., correct answers. In a six-answer test the median was 5, and in an eight-answer test it was 7.

The percentages of wrong answers by topics in both tests were:

Factoring		13.1
Difference of squares	14.7	
Trinomial square	10.3	
x^2+ax+b	14.3	
Fractions		28.0*
Addition	62.2	
Multipl'cation	26.6	
Division	29.3	
Linear equations		29.2
Integral	26.4	
Fractional	35.4	
Literal	49.4	
Simultaneous	21.5	

* Computed without the addition question, which was recognized as too difficult.

The addition question was a little too long. Inferences were drawn as to the relative difficulty of the different topics. Regarding factoring and literal equations conclusions were reached in harmony with the recommendations of the National Committee on Mathematical Requirements.[1]

FOURTH-SEMESTER ALGEBRA

A general test was given to the seventeen classes in the last semester of algebra taught. The simple averages for all classes and schools were computed, tabulated, and ranked. The average marks by topics in fourth-semester algebra were as follows:

Solving quadratics......................	86.3
Simplifying radicals.....................	51.2
Problem involving quadratics:	
Equation correctly formed.............	15.7
Complete solution....................	10.3
Drawing graph.........................	15.5
Finding square root.....................	67.2

The test was too long, so that only limited conclusions could be drawn regarding the later questions, but practically all of the pupils had time to finish the questions on quadratics and on radicals.

The average for solving quadratic equations appeared very high and was the result of extensive practice.

The understanding and manipulation of powers and roots represent a rather advanced type of mathematical thinking. As 50 per cent success is not of much value, the results in radicals seemed to indicate the desirability of limiting the amount of ground covered in the field of radicals and exponents and increasing the practice in the more restricted field until greater precision is secured. This is in harmony with the recommendations of the Commission on the Reorganization of Secondary Education[2] and the National Committee on Mathematical Requirements.[3]

[1] *The Reorganization of Mathematics in Secondary Education*, pp. 20 and 23, Bureau of Education Bulletin No. 32, 1921.

[2] *The Problem of Mathematics in Secondary Education*, p. 19. Bureau of Education Bulletin No. 1, 1920.

[3] *The Reorganization of Mathematics in Secondary Education*, p. 20. Bureau of Education Bulletin No. 32, 1921.

The problem involving the use of quadratics was rather difficult; in fact, it is hard to find problems involving quadratics which are not difficult. One might add that real problems are rare. While simple equations are applicable to numerous situations in everyday life, quadratic equations have few practical applications. Only 15 per cent of the pupils had sufficient insight into the problem to form a correct equation, and only 10 per cent reached a final solution. Of course, these percentages would have been somewhat higher if all of the pupils had reached this problem. Nevertheless, the meager results in the *use* of quadratic equations, together with the fact that any real uses of them are hard to find, seemed to throw doubt on the justification of spending such an extensive amount of high-school time on quadratics as at present.

The question on graphs revealed the fact that, though graphs are theoretically a part of our course of study, the teachers are not in general teaching them. The advanced text adopted by the state is out of date, and practice in making graphs enters in only on the initiative of the teacher. The result is that up to this time we have not been doing justice to this very practical and important phase of the subject. Graphic treatment and interpretation should be prominent throughout the algebra course.

Over two-thirds of the pupils produced correct results in extracting square root, which is a good record when we take into account the fact that not all of them had time to reach this question.

The performance on separate topics was also tabulated for the separate schools and the separate classes, showing just which teachers were teaching graphs and the high and low performances of each class and on each topic.

One rough measure of a pupil's ability may be found in the average of the marks which he receives in all of his subjects. These averages were found and brought into comparison with the averages on the test. The ratio of test average to the ability index was computed, and this ratio then formed what McCall would designate as an accomplishment quotient,[1] that is, the amount of accomplishment in proportion to ability. Classes were then ranked both

---

[1] W. A. McCall, *How to Measure in Education*, pp. 85–87. New York: Macmillan Co., 1922.

ways, by their simple "performances" and by their "accomplishments" in proportion to ability. It was found that eleven classes remained unchanged in rank; four were displaced one rank; two were displaced two ranks, and the average displacement was only one-half of a rank. That is, it made no appreciable difference in appraising these classes whether we used an "accomplishment" ratio or the simple "performance" mark.

The question was asked in another form, whether the ranks the classes attained on the test merely reproduced their ability ranks or whether due to teaching the test ranks proved to be different from the ability ranks. It appeared that the classes by no means arranged themselves in test ranks in the order that would be expected from their ability indexes. Only one class had the same rank in both, the others varying from one to ten places, with an average discrepancy of 4.4 places.

The test results were compared in detail with the teachers' final examination marks and with the term marks. Inferences were drawn as to variations among teachers with regard to the difficulty of their final examinations and the extent to which the teachers were too high or too low in general marking.

LAST SEMESTER OF PLANE GEOMETRY

A test was given to the ten classes completing plane geometry. The usual instruction in geometry and the usual testing consist largely of reproducing demonstrations which are printed in the textbook. In this test all questions were in the nature of "originals" or "applications" of principles learned. For this reason the test was very much harder than an ordinary test; it was given in this form to ascertain the amount of ability being developed to deal with original situations and to make applications of principles learned. The test proved to be too long for the single period in which it was given.

The test proper consisted of three questions. The first required the demonstration or proof of a theorem. The second required a geometrical construction. The third involved the practical application of a known principle. A fourth, a problem in computation, was added for any who might be able to do more in the time allowed than the regular test called for.

The general distribution of marks and the averages for the separate classes and schools were worked out. It was found that the amount of ability being developed by pupils in this subject varied greatly.

Marks on different phases of work were computed for each class and school and for the city. The city averages on the different phases of the work follow. It should be borne in mind that pupils who did not reach a given question were scored as zero.

    1. Demonstration of "original" theorem.............. 61.3
    2. Problem in construction.......................... 35.5
    3. Application of principle.......................... 38.8
    4. Application involving computation................ 8.0

Only a very few reached the fourth question. The best score was made on the demonstration of a theorem, and this was the type of work on which they had had the most practice. How much higher they would have averaged on the demonstration of a book theorem instead of an original was not determined by this test. There was some ability to proceed independently with theorems similar to those in the book, but this ability was not as highly developed as is desirable.

There was less power to deal with problems of construction than with the demonstration of theorems.

The ability to apply principles already learned was poor. For this the classroom teachers cannot be held wholly responsible, as the number of book theorems now required to be taught is too large. Apparently, the number of book theorems to be learned should be reduced in order to secure more time for practice in applications. This statement is in harmony with the report of the National Committee on Mathematical Requirements.[1]

Test averages and the pupils' ability were compared as in the case of the more advanced algebra test. In this instance for ten classes the test ranks differed from the ability ranks by an average of 2.2 places. The chance difference would be an average of 4.6 places. Consequently we have here indicated clear connection between the two, though by no means complete agreement.

[1] *The Reorganization of Mathematics in Secondary Education*, chap. vi.   Bureau of Education Bulletin No. 32, 1921.

The papers of this test were first marked by the several teachers in accordance with general directions; later they were marked by the supervisor in a somewhat simpler manner. The average for the city according to the supervisor's marks was within three-tenths of 1 per cent of the average according to the teachers' marks. With the supervisor's average mark as a uniform standard it was then ascertained how much too high or too low each teacher and school appeared to mark.

<div align="center">FIRST-SEMESTER LATIN</div>

Ten classes in beginning Latin took a test on the forms of the language studied up to that time.

The simple averages and averages on a standardized scale were computed for each class, and the general distribution of the pupils' marks was ascertained.

The percentage of errors was calculated for every separate form given in the test. The percentage of wrong forms in the second declension was 7; in third declension consonant stems, 13; and in third declension *i*-stems, 16, the average being 12.

In the verb forms 26 per cent were wrong in the first conjugation and 31 per cent in the second, or an average of 29 per cent wrong in all of the conjugation work given.

From these figures it was evident that the proportion of error on verb forms is very much higher than on nouns, the amount of error being nearly 30 per cent on verbs and only 12 per cent on nouns. In view of this, and taking into account the fact that there are only fifty or sixty standard noun forms, while the standard verb forms number six or seven hundred, it was inferred that it would be wise to memorize only one-third of the verb forms in the first year, namely, those of the third person; especially as the other persons are not used in the text of Caesar and there is no need of employing them in reading until Cicero is reached in the third year. It seemed that a pupil with thorough mastery of the third person would be able to acquire the first and second persons readily when he came to study Cicero. Since this policy has been adopted the teachers report improved grasp of inflections in the first year.

SIXTH-SEMESTER LATIN

A test was given to the three classes just completing their third year of Latin; this is the point at which they are doing their last high-school reading in Latin prose literature, the next year being devoted to Latin poetry. It was desired to ascertain how much power to read Latin prose literature is being developed in the high-school course.

In the time allowed, most of the pupils were able to answer only the first three questions, of which the first was a passage for translation, the second a series of questions designed to test the extent to which the pupil understood the significance of what he was reading, and the third required an exact statement of the inflectional forms of certain words. The passage for translation was somewhat overdifficult. The questions on the significance of what was read were very searching.

A few pupils had time to write answers to four other questions, one on syntax, one on derivatives, and two dealing with the relation of Latin to English.

The averages made on the separate questions by those pupils who had time to answer them, each average being expressed on a 100-point scale, are shown in Table I.

TABLE I

	Class 1	Class 2	Class 3	All Classes
1. Translation....................	52.6	34.0	68.3	55.2
2. Understanding and knowledge....	48.0	56.7	52.0	52.0
3. Inflections...................	44.0	26.0	30.0	33.0
4. Syntax......................	51.0	....	39.0	43.0
5. Derivatives..................	76.0	....	85.0	82.0
6. Relation of Latin to English......	90.0	....	96.0	94.0
7. Relation of Latin to English......	....	....	44.0	44.0

The test was primarily one in translation. The passage selected for translation had been read in the early part of the term. A more recent selection was not taken because the different schools were not reading exactly the same things at the end of the term, this not being required by the course of study. All had read, though some time previously, the oration from which the selection

was taken.   Marking was rigidly mathematical, giving credit for just such fractions of the whole as were right.

It might seem, when a careful marking showed the three schools reaching only 68 per cent, 53 per cent, and 34 per cent correct translation, or an average for the city of 55 per cent, that not much is being accomplished in Latin.   But this depends somewhat on the criterion of judgment.   We must remember that in mathematics, science, etc., it is theoretically possible to select and adapt the curriculum material to the degree of maturity of the pupils. In Latin, however, we do not bring the pupil in contact with Latin text especially prepared for him and exactly suited to his degree of maturity.   Instead, the fundamental theory in Latin requires that we shall use real Latin text, written two thousand years ago for people whose native tongue was Latin.   Cicero is used not because the text is easy but primarily because Cicero was the greatest Latin writer.

Even if a similar test should be given in a modern foreign language, which in many ways would be much easier, we should be disappointed if we expected a rigid marking to show any close approach to 100 per cent efficiency and accuracy.

That no such high scores can be expected or attained in Latin classes as ordinarily conducted can be seen by reference to a study[1] by H. A. Brown of the Latin students in fifteen high schools in New Hampshire.   Brown made the most extensive study of the results of Latin instruction in high schools which has yet been attempted, though it will be surpassed by the investigations now projected by the Classical League.   One of Brown's tests was in the translation of connected Latin.   The Brown test differed from the Dallas test in the fact that in the former the passage was sight reading, while in the latter the passage had been read a couple of months previous to the test.   On the other hand, the Brown test used a passage from Caesar which was distinctly easier than the Cicero passage.   The marking in the two cases was comparable though not exactly the same.   In comprehension of meaning, as shown by translation, the third-year pupils of the fifteen high

[1] H. A. Brown, *A Study of Ability in Latin in Secondary Schools*, chap. vi.   Oshkosh, Wisconsin: State Normal School, 1919.

schools in New Hampshire with the largest enrolments ranged from
21 per cent to 68 per cent and averaged 45 per cent as compared
with a range of from 34 per cent to 68 per cent and an average of
55 per cent for the city of Dallas.   So far as we can judge, the
results here are substantially similar to those in the larger high
schools of New Hampshire.

Other aspects.—The second question was intended, independently
of translation, to test the pupil's understanding and knowledge of
the situation involved in the passage read.   The scores here would
have been higher if the class reading had occurred more recently.
The marks indicated that the pupils' success in grasping the his-
torical setting and situation was just about the same as their success
in translation, though the separate classes varied.

The marks on the "inflections" question seem very low, but
this is mainly due to incompleteness rather than error.   Errors
were not very numerous, but the pupils often failed to make a
complete statement of all of the details implied in the form of the
question.   The answers were marked rigidly on the basis of com-
pleteness of statement.

The syntax question was marked with reference to grasp of the
essential idea rather than completeness of statement, and the
scores were somewhat higher than for inflections.

The questions on English derivatives and on the relation of
Latin to English were easy and the scores relatively high.   There
was no indication that the relation of Latin to English was being
neglected, the lowest scores being in those questions dealing
essentially with the Latin itself.

One of the problems to be considered by the Classical League
in its investigation is whether the traditional objective of power to
translate classical Latin shall remain the central objective or
whether it shall be abandoned in favor of some objective which
takes more account of the degree of maturity of the pupils and
the probable brevity of their Latin course.   Results such as those
from this test have bearing on the question.

Comparisons were also made of the test marks with the teachers'
final examination and term marks and with ability indexes formed
by computing each pupil's average in all subjects.   The class

making the lowest score on the test was also lowest in ability.   The
ranking in test performance was the same as the ranking in "accomplishment" in terms of ability (that is, the ratio of test performance
to ability index).

SECOND-SEMESTER HISTORY

A test was given to twenty second-semester classes studying
medieval history.   Simple averages and averages on a standardized
scale were worked out and compared with the teachers' marks.

The test consisted of information questions of fact or identification.

The percentages of wrong answers for the various types of
information questions were:

Fact of social custom....................	8.0
Fact of event...........................	26.8
Identification of person..................	30.4
Non-personal identification...............	28.8
Identification of place...................	27.6